

**DTIC FILE COPY**

2

Document No. 1800-001  
1 July 1990

AD-A228 476

**SGML Product Review  
for the  
Software Technology for Adaptable, Reliable Systems  
(STARS) Program**

**Contract No. F19628-88-D-0032**

**Task 1R65 - SGML Document Descriptions**

**CDRL Sequence No. 1800-001**

1 July 1990

**DTIC**  
**ELECTE**  
**NOV 09 1990**  
**S B D**

**Prepared for:**

**Electronic Systems Division  
Air Force Systems Command, USAF  
Hanscom AFB, MA 01731-5000**

**Prepared by:**

**IBM Federal Sector Division  
800 North Frederick Avenue  
Gaithersburg, MD 20879**

Links on Distribution: U.S. Government  
Agencies and their Contractors  
Requests to DoD Controlling Office

**DISTRIBUTION STATEMENT A**  
**Approved for public release**  
**Distribution Unlimited**

## REPORT DOCUMENTATION PAGE

Form Approved  
OMB No 0704-0188

1. The information contained in this document is confidential and is intended for the use of the individual named in the title of this document. It is not to be distributed outside the organization without the express written consent of the individual named in the title of this document. It is not to be used for any purpose other than that for which it was prepared. It is not to be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage and retrieval system, without the prior written permission of the individual named in the title of this document. It is not to be used for any purpose other than that for which it was prepared. It is not to be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage and retrieval system, without the prior written permission of the individual named in the title of this document.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE July , 1990		3. REPORT TYPE AND DATES COVERED Final	
4. TITLE AND SUBTITLE SGML Product Review				5. FUNDING NUMBERS C: F19628-88-D-0032	
6. AUTHOR(S) S. Kutoroff					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  IBM Federal Sector Division 800 N. Frederick Avenue Gaithersburg, MD 20879				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  Electronic Systems Division Air Force Systems Command, USAF Hanscom AFB, MA 01731-5000				10. SPONSORING / MONITORING AGENCY REPORT NUMBER  CDRL Sequence No. 1800- 001	
11. SUPPLEMENTARY NOTES					
12a. DISTRIBUTION / AVAILABILITY STATEMENT				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  SGML (Standard Generalized Markup Language) is an international standard for representing the elements and structure of electronically stored text.  This document reviews commercial SGML products.					
14. SUBJECT TERMS  STARS, SGML, Standard Generalized Markup Language				15. NUMBER OF PAGES 32	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT  Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE  Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT  Unclassified	20. LIMITATION OF ABSTRACT  UL		

## TABLE OF CONTENTS

1. SGML Products	1
2. SGML Software Vendors	5
3. AGFA Compugraphic Division	6
4. ArborText, Inc.	7
5. Avalanche Development Company	9
6. Datalogics Incorporated	11
7. IBM Corporation	15
8. SoftQuad Inc.	18
9. Software Exoterica Corporation	21
10. Yard USA Inc.	25
A. APPENDIX: Product Types by Vendor	27
B. APPENDIX: Products and Product Types	28
C. APPENDIX: Products and Platforms	29
D. APPENDIX: Products by Vendors	30



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By <i>per letter</i>	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
<i>A-1</i>	

## 1. SGML Products

This document details the software products currently available for creating and processing documents using the Standard Generalized Markup Language (SGML) as specified in ISO 8879:1986 and in MIL-STD-28001A (CALS).

### 1.1 SGML Product Types

SGML products range from complete systems (which do everything from allowing a user to create a document to printing a formatted document) to smaller products which are used for only one or two specific functions in the SGML document processing model. SGML document processing involves the following functions:

- 1.) creating a Document Type Declaration (DTD)
- 2.) parsing a DTD
- 3.) creating a Document Instance (document marked-up in SGML based on DTD)
- 4.) parsing a Document Instance
- 5.) translating the SGML in a Document Instance into a formatting language for output

The advantage of using a product which performs only a few of these functions lies in the price and in the ability to customize the products to the user's needs. This section will describe SGML product types and the function each performs in producing a SGML document.

#### 1.1.1 Parser

A parser validates the mark-up of a SGML document based on a DTD which has either been system defined or compiled with a DTD Compiler. A parser which is fully compliant with ISO 8879 follows all SGML syntax rules including parsing SGML declarations, SGML DTDs and SGML source document instances. The parser detects and reports all violations of the SGML standard. Many products advertised as ISO 8879 compliant parsers only work with a few vendor supplied DTDs and require an additional product (and additional money) to parse and use DTDs designed by the user. To distinguish between those parsers which can use any DTD versus those that can only work with vendor supplied DTDs, this document will use the term "SGML parser" for a SGML parser which parses a source document instance based on DTDs supplied by the vendor. A parser that can also parse DTDs will be referred to as a "SGML/DTD parser".

#### 1.1.2 DTD Compiler

A DTD compiler is a product which allows the user to create custom document types for use with a SGML parser and/or DTD Based SGML Editor. The DTD compiler creates a file or library entry that is in turn used by a SGML parser/editor. A DTD compiler is required if the user wishes to design documents and does not have a SGML/DTD parser.

#### 1.1.3 Normalizer

A normalizer is a SGML parser which accepts SGML minimized input (using OMITTAG feature) and creates as output a fully marked-up SGML document (all front and end tags included). This type of product is useful if the user is creating SGML documents which may be transmitted to another company which does not have a SGML system which handles minimized mark-up. It is also useful if SGML output is being used to tag data for loading into a database.

#### 1.1.4 Translator

A translator is a product which inserts SGML mark-up into non-SGML documents and/or translates SGML mark-up into a document containing formatting commands for output, insertion into a database, or other applications. A translator is a necessary step in the SGML process if the user is not using a full SGML system that handles document output. A translator allows the user to define an output format for a SGML DTD and translate a SGML document into an output format that is acceptable for the user's need.

#### 1.1.5 Tagger

A tagger adds SGML mark-up to non-SGML documents. A translator can often perform the same tasks as a tagger, but a tagger will usually only insert SGML mark-up and not formatting instructions. A tagger may take input from a word processor, text file, scanned document, etc. A tagger is used when transferring from one way of producing documents, such as traditional word processors, to SGML. The tagger allows established document creation procedures to continue while complying with the SGML standard.

#### 1.1.6 DTD Based SGML Editor

A DTD Based SGML Editor is an editor which allows the user to create a SGML document by prompting for valid SGML tags and/or validating mark-up inserted into a document. This editor will usually include a SGML parser, but often requires a DTD compiler to be purchased separately if the user wishes to use custom document types.

#### 1.1.7 Application Generator

An application generator is usually a library of sub-routines that can be incorporated into customized applications developed by the user. Generally, the sub-routines are calls to a SGML parser that can be used to develop custom editors, database applications, etc.

#### 1.1.8 Host Interface

SGML products which are based on mainframes or minicomputers often have a PC or Workstation based product that can be used as an interface to the main system.

#### 1.1.9 Formatter

A formatter is a product which produces a printed document from a SGML marked-up document and/or a formatting language style sheet. The main difference between a translator and a formatter is that the translator usually outputs a document in a formatting language which then must be processed with a formatter to get a printed document.

#### 1.1.10 Graphics System

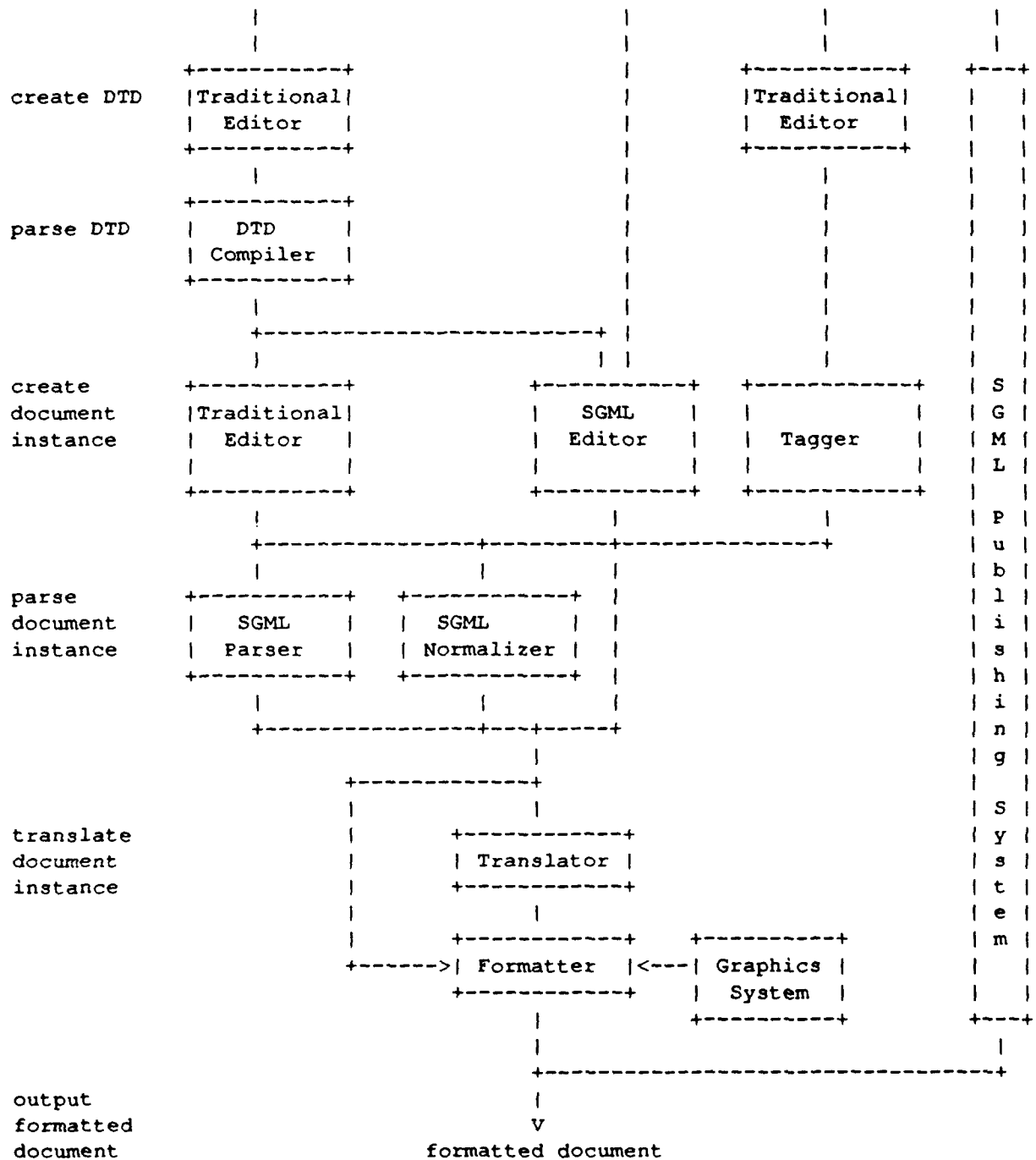
A graphics system is a product which can incorporate graphic displays/drawings into a SGML document during formatting. Graphics are stored in standard formats such as Computer Graphics Metafile (CGM).

#### 1.1.11 SGML Publishing System

A SGML publishing system encompasses most if not all of the previously mentioned product types. Using a SGML publishing system an author can create a SGML document (via an

editor) and translate it to a medium which can be printed (translator, formatter, graphics system). The most recent trend in SGML publishing systems is to incorporate an object oriented database into the SGML system. This database stores document elements to be reused in several documents. This is especially helpful for graphics, tables, standard warnings, etc.

The following figure summarizes the types of SGML products available and indicates how these product types relate to each other:



## 2. SGML Software Vendors

The following sections present some of the vendors for available SGML products. Vendors are presented alphabetically. Each section gives the name and address of a single vendor as well as information on each product. This information includes product types (described previously), a brief description of the product, supported platforms, and a sampling of prices for selected platforms. Product type information is divided into a primary product type and secondary product types. Secondary product types are used when a product could actually be used in place of one or more other products. Platform and price information are meant to be used as guidelines and not as an inclusive list. Many vendors support platforms other than those listed and price information is constantly changing. Readers of this document are advised to contact the vendor for further information.

NOTE: This document uses product names which are trademarks of their respective vendors. Use of these trademarks is for editorial reasons only with no intent of trademark infringement.



### 3. AGFA Compugraphic Division

AGFA Compugraphic Division specializes in large publishing and typesetting systems. They currently offer one SGML product, Compugraphic's Automated Publishing System (CAPS) which is a full SGML System.

AGFA Compugraphic Division  
200 Ballardvale Street  
Wilmington, MA 01887  
Phone: 508-658-5600  
Fax: 508-657-7160

#### 3.1 Compugraphic's Automated Publishing System (CAPS)

Product Class: SGML Publishing System

CAPS is a full SGML publishing system integrating SoftQuad's Author/Editor with a full technical publishing system. CAPS was designed specifically for SGML and uses SGML as a data model from document creation through formatting. Use of custom DTDs (DTD compiling) is supported. All CALS publishing standards are supported. In addition to the DTD Based SGML Editor, input may be received from traditional text editors and word processors as well as several scanners and graphic processors. Formatting features include batch pagination, automatic kerning, white space control, widow and orphan control, automatic table placement, cross-referencing, indexing, table of contents, automatic numbering of sections, paragraphs, tables, etc. Output is to Postscript laser printers, imagesetters or 1840A tape. CAPS supports full revision control including multiple levels, displayed or suppressed revisions and the output of change pages. Future enhancements to the CAPS system will include the integration of an object oriented database.

Platforms: SUN

Price:

- o based on application and volume
- o \$15,000 for System Management Software (per site)
- o \$12,500 for Base CAPS Application (per site)

#### 4. ArborText, Inc.

ArborText, Inc. specializes in publishing and typesetting systems. They currently offer one SGML product, The Publisher, which is a full SGML System.

ArborText, Inc.  
553 West William Street, Suite 300  
Ann Arbor, MI 48103  
Phone: 313-996-3566  
Fax: 313-996-3573

##### 4.1 The Publisher

###### Product Class: SGML Publishing System

The Publisher is a full publishing system based on SGML. This system supports multiple windows which can simultaneously show the document being edited and a graphic representation of the final document. The Publisher features CALS and SGML support, customizable keyboard and menus, X-Window support, text filters for several input formats, a table editor, Hypertext links within and between documents, a spell checker, a thesaurus, a dictionary, automatic display of encapsulated Postscript files, table of contents for section, chapter or document, landscape option within a document, incorporation of Mathematica(tm) results, relative font sizing, mail merge feature, shaded boxes and text, paragraph numbering, document locking, autosave and extensive graphics support including an equations editor. The Publisher will have a new CALS release later this year that will include the ability to use custom DTDs and have the ability to integrate a database application with The Publisher.

###### Platforms:

- o SUN 3 & 4, Sun386i, Sun SPARCstation
- o Apollo Series 2500, 3500, 4500
- o HP 9000 series 300

###### Price:

###### License Fee

- o \$3495 for 1 license
- o \$2995 for 2 - 9 licenses
- o \$2495 for 10 - 29 licenses

###### Annual Maintenance Fee

- o \$525 for 1 license
- o \$450 for 2 - 9 licenses
- o \$375 for 10 - 29 licenses

Annual Maintenance Fee with Graphics Packages

- o \$600 for 1 license
- o \$525 for 2 - 9 licenses
- o \$415 for 10 - 29 licenses

Import filters, PubPaint & PubDraw Graphics packages, HP LaserJet Driver, Postscript Documentation Files, Training and Consulting extra - contact vendor.

## 5. Avalanche Development Company

Avalanche Development Company specializes in data preparation software including SGML tagging products. They currently have two products on the market that do SGML tagging.

Avalanche Development Company  
947 Walnut Street  
Boulder, CO 80302 USA  
Phone: 303-449-5032  
Fax: 303-449-3246

### 5.1 FastTag

Product Class: Tagger

FastTag allows translation of non-SGML files including ASCII, print image files (Epson, Diablo, IBM), OCR scanner files, and wordprocessor files to SGML markup or other output formats such as XICS, troff, Datalogics, SoftQuad, etc.

Platforms:

- o IBM PC/AT and compatible with 640K (MS-DOS)
- o SUN 3 & 4 (SUN OS 3.5 & 4.0)
- o Apollo (SR10.0)
- o Macintosh

Price:

- o \$2,450 for IBM PC/AT and compatibles, and Macintosh
- o \$3,100 for SUN and Apollo

#### 5.1.1 FastTag CALS Application package

The FastTag CALS Application package allows FastTag to translate documents into CALS SGML.

Price:

- o \$500 for IBM PC/AT and Macintosh
- o \$750 for SUN and Apollo

### 5.2 IMSYS.CALS

Product Class: Tagger

IMSYS.CALS allows translation of non-SGML files including ASCII, print image, scanner, and wordprocessor files to CALS SGML. Similar to FastTag with CALS Application

package, but IMSYS.CALS produces only CALS output and is production oriented with a conversion rate of 2-3 seconds per page.

Platforms:

- o IBM PC/AT and 100% compatibles with 640K (MS-DOS)
- o SUN 3 (SUN OS 4.0)
- o Apollo DN-3000 (SR10.0)
- o MacIntosh

Price:

- o \$2,450 for IBM PC/AT and compatibles
- o \$3,100 for SUN and Apollo

## 6. Datalogics Incorporated

Datalogics has a wide assortment of SGML products as well as other full composition systems, translators and text manipulation products. Only those products recommended for SGML applications are listed here. These products can be combined to create a complete SGML Publishing System (Datalogics Technical Publishing Environment).

Datalogics Incorporated  
441 West Huron Street  
Chicago, IL 60610-3674 USA  
Phone: 312-266-4444  
Fax: 312-266-4473

### 6.1 ParseStation

Product Class: SGML Parser

Secondary Classes: DTD Compiler, Normalizer

ParseStation is a complete SGML Parser that can operate either in menu or batch mode. ParseStation features DTD compiling, and optional normalization of SGML documents. ParseStation supports all features specified for CALS SGML.

Platforms:

- o DEC VAX (VMS)
- o IBM-compatible XT or AT (MS-DOS), OS/2
- o Xenix
- o DECstation 3100 (Ultrix)

Price:

- o \$3,500 for VAX 3100, \$5,000 for VAX 3600
- o other prices available from vendor

#### 6.1.1 ParseStation C library

Product Class: Application Generator

Secondary Classes: SGML Parser, DTD Compiler, Normalizer

The ParseStation C library is a library of C subroutines for the ParseStation. These subroutines can be used to customize applications.

Platforms:

- o DEC VAX (VMS)
- o IBM-compatible XT or AT (MS-DOS), OS/2
- o Xenix
- o DECstation 3100 (Ultrix)

Price:

- o \$3,500 for VAX 3100, \$5,000 for VAX 3600
- o other prices available from vendor

## 6.2 DL Pager

Product Class: Formatter

The DL Pager is an automated, high-speed, batch pagination and composition system for high-volume publishing. DL Pager accepts SGML markup as input and supports several output devices including Postscript laser printers and typesetters, APS photosetters, Compugraphics photosetters, and Datalogics 6380 laser printers. Features include hyphenation, justification, user-specified kern pairs, aesthetic ragged text, automatic skewing of indents, fractions, small caps, floating accents, widow/orphan control, vertical justification, automated table of contents and indices, graphic space reservation, sidemarks/change bar insertion, single pass layout, multiple font capabilities, and typesetting customization. To include graphics in output composed by DL Pager, the Datalogics Integrated Graphic System must be used.

Platforms:

- o DEC VAX (VMS)

Price:

- o \$110,000 for VAX 3600
- o other prices available from vendor

NOTE: A output module must be purchased with DL Pager. Cost is \$10,000-15,000 for a VAX 3600.

### 6.2.1 PageStation

Product Class: Host Interface

An interface to DL Pager that allows display of formatted text in a terminal window.

Platforms:

- o VAX Workstation

Price:

- o \$10,000 for VAX 3600
- o other prices available from vendor

#### 6.2.2 PC-Proof

**Product Class: Host Interface**

An interface to DL Pager that allows display of formatted text in a terminal window.

**Platforms:**

- o contact vendor

**Price:**

- o contact vendor

#### 6.3 Integrated Graphic System (IGS)

**Product Class: Graphics**

The Integrated Graphic System manages and manipulates graphics. The IGS accepts several input formats including Postscript, Hewlett-Packard Graphics Language, IGES, Interleaf, Lotus 1-2-3 PIC, Macintosh MacPaint, Macintosh PICT I, PC Paintbrush PCX, Textet Image Format, TIFF, Xerox 9700 IMG, DEC Compound Document Architecture, and DEC Document Interchange Format. Input graphics are converted for output or for inclusion in document being processed by DL Pager. Output formats include APS photosetter, CCITT Group 4, Datalogics laser printer, Datalogics PC-Proof, Datalogics PageStation, Datalogics DECwindows, Lasercomp typesetters, Postscript laser printers and typesetters, and Xerox 9700 IMG laser printer. IGS is primarily a graphic format translator; other options include graphic rotation and resizing.

**Platforms:**

- o DEC VAX (VMS)

**Price:**

- o \$10,000-15,000 for graphic output
- o \$15,000 for graphic descriptor file manager
- o \$ 5,000-10,000 for graphic import

(all quotes for VAX 3600, other prices available from vendor)

#### 6.4 WriterStation

**Product Class: DTD Based SGML Editor**

**Secondary Classes: Host Interface**

The WriterStation is a DTD based SGML editor. It can work as a stand-alone editor on a



PC or as a PC interface to a VAX minicomputer. If used as a PC interface, WriterStation may be integrated with DL Pager for full publication functions. The WriterStation features multiple files and windows, customization of editor functions and help screens, menu tag prompting, viewing all occurrences of a specific tag, on-screen formatting, markup validation, an on-line spell-checker and thesaurus.

Platforms:

- o IBM-compatible PC

Price:

- o \$1,500 for IBM-compatible PC

#### 6.4.1 CALS Application package

The CALS Application package supports documents that conform to CALS standards.

Price:

- o \$200-\$300

#### 6.4.2 WriterStation Utilities

Product Class: DTD Compiler

The WriterStation Utilities allow any DTD to be used with WriterStation. Additional functionality includes the ability to define screen formatting rules for all markup in a document type, define generated text and create multi-column screen display and print layouts, define enumeration rules for elements, create menus for valid markup and attribute data, and use search/replace operations on elements.

Price:

- o \$6,000

## 7. IBM Corporation

IBM developed the generalized markup language GML which was the forerunner of SGML. IBM has a complete line of SGML products for IBM mainframes, workstations and PCs. Only SGML specific products are listed here. When combined, these products form a full SGML Publishing System.

IBM Corporation  
Publishing Systems Information Development  
Dept. G81  
P.O. Box 1900  
Boulder, CO 80301-9191 USA  
Phone: contact local IBM representative

### 7.1 SGML Translator DCF Edition

Product Class: Translator

Secondary Classes: SGML Parser

SGML Translator DCF Edition validates and translates SGML to a format that can be used by IBM's Document Composition Facility (DCF). Included with the product is the DTD for CALS MIL-M-28001.

Platforms:

- o IBM host (VM or MVS)

Price:

- o \$20,790 for IBM S/370

### 7.2 SGML TextWrite OS/2 Edition

Product Class: DTD Based SGML Editor

SGML TextWrite OS/2 Edition is a SGML-smart editor. Features include markup prompting, validating, on-screen formatting and printing of a draft copy at an attached printer. The SGML TextWrite Editor creates and validates CALS compliant documents; in order to customize applications, the SGML TextWrite Tools must be used.

Platforms:

- o IBM PS/2 (OS/2)

Price:

- o \$1,650

### 7.2.1 SGML TextWrite Tools OS/2 Edition

Product Class: DTD Compiler

SGML TextWrite Tools OS/2 Edition allows any DTD to be used with with SGML TextWrite OS/2 Edition. This product validates the DTDs, creates menus for the new DTD tags and enables a style sheet for the new DTDs to be created.

Platforms:

- o IBM PS/2 (OS/2)

Price:

- o \$5,500

### 7.3 TextTagger Workstation Edition

Product Class: Translator

TextTagger Workstation Edition translates untagged document files to CALS SGML mark-up. TextTagger accepts files from several word processors, printers, and scanners.

Platforms:

- o IBM PS/2 (OS/2 or PC-DOS)

Price:

- o \$5,445

### 7.4 ProcessMaster TextTagger Feature

Product Class: Translator

ProcessMaster TextTagger Feature translates untagged document files to CALS SGML mark-up. ProcessMaster TextTagger Feature accepts files from several word processors, printers, scanners. ProcessMaster TextTagger Feature differs from TextTagger in that it allows for other input (such as formatted text) and output (such as BookMarker and SGML starter set mark-up).

Platforms:

- o IBM host (VM or MVS)

Price:

- o \$11,160 for IBM S/370

## 7.5 ProcessMaster CALS Application Feature

Product Class: Formatter

Secondary Classes: Graphic System

Allows Document Composition Facility to format documents according to MIL-M-38784B specifications. Allows inclusion and management of graphic files as well as the writing of CALS documents to CALS compliant 1840A tapes. Manages CALS-compliant tapes. One version of the ProcessMaster CALS OS/2 Workstation Driver Feature is included with this system.

Platforms:

- o IBM host (VM or MVS)

Price:

- o \$29,500 for IBM S/370

### 7.5.1 ProcessMaster CALS OS/2 Workstation Driver Feature

Product Class: Host Interface/Driver

The Workstation Driver is a host interface that includes editing text, tagging text, parsing SGML mark-up, creating/editing/transforming graphics and images, printing drafts and final copy at the host, printing a draft copy at the workstation, transferring files between the workstation and the host, managing files on host libraries, and reading and writing to tapes. One copy of this product is included as part of the ProcessMaster CALS Application Feature.

Platforms:

- o PS/2 (OS/2) cooperating with a host machine (VM or MVS)

Price:

- o \$139

## 8. SoftQuad Inc.

SoftQuad offers a full publishing line including SGML parsers, SGML editors, SGML taggers, a publishing system, and X-Windows support.

SoftQuad Inc.  
720 Spadina Avenue  
Toronto, Canada M5S 2T9  
Phone: 416-963-8337  
Fax:

### 8.1 Author/Editor

Product Class: DTD Based SGML Editor

Author/Editor is a SGML-smart editor. Features include markup prompting, validating, on-screen formatting, inclusion of graphic material, formatting including handling of tabular formats, and printing of a formatted copy. Tags can be entered from a menu or from the keyboard. Author/Editor comes with several DTDs. In order to create new document types, the RulesBuilder product must be purchased.

Platforms:

- o Macintosh Plus, Macintosh SE, Macintosh II (including A/UX)
- o Sun 3 & 4

Price:

- o \$495 for Macintosh
- o \$995-\$1,695 for Sun

#### 8.1.1 Author/Editor CALS Version

Product Class: DTD Based SGML Editor

Author/Editor CALS Version is set up to work with the MIL-M-28001A DTD.

Platforms:

- o Macintosh Plus, Macintosh SE, Macintosh II (including A/UX)
- o Sun 3 & 4

Price:

- o \$995 for Macintosh
- o \$1,295-2,595 for Sun

### 8.1.2 RulesBuilder

Product Class: DTD compiler

RulesBuilder allows any DTD to be used with Author/Editor.

Platforms:

- o Macintosh Plus, Macintosh SE, Macintosh II (including A/UX)
- o Sun 3 & 4

Price:

- o \$995 for Macintosh

### 8.2 MARK-IT, the Sobemap Parser

Product Class: SGML Parser

SoftQuad is one of two distributors of MARK-IT listed in this document. MARK-IT is a complete SGML Parser that operates in batch mode. MARK-IT features DTD compiling, error recovery, entity declaration updating within document instances, counters to identify control and sequence of elements, creation and handling of linked document structures, linking documents to non-SGML applications such as typesetters, retroactive tagging and complete conformance to ISO 8879:1986.

Platforms:

- o Standard PC, 386 or PS/2
- o workstations/mini computers/mainframes (1- 40 terminals) (VAX-VMS,UNIX,MS-DOS)

Price:

Basic Version:

- o \$ 1,000 for Standard PC
- o \$ 2,000 for UNIX or PC Network(up to 8 seats)
- o \$ 6,000 for mainframes, mini & micro computers

Full Version:

- o \$ 5,000 for Standard PC
- o \$10,000 for UNIX or PC Network(up to 8 seats)
- o \$30,000 for mainframes, mini & micro computers

Site Licenses:

- o \$125,000 for unlimited use on single platform

- o \$160,000 for unlimited use on multiple platforms

### 8.3 FastTag

Product Class: Tagger

SoftQuad is a distributor of FastTag which was previously described under Avalanche Development Company.

### 8.4 SoftQuad Publishing Software

Product Class: Formatter

SoftQuad Publishing Software is an automated production publishing system. Features include table formatting, mathematical formula and equation formatting, graphics, and charts. Other features include justification, centering hyphenation, page and section numbering, multiple columns, kerning, user-definable hyphenation dictionary, conditional formatting and pagination, vertical justification, widow control, a readable intermediate language that can be modified using awk and grep, and a diagnostic and debugging system. Input is accepted from several word processors and editors including Author/Editor. Output devices include laser printers, laser printers and typesetters that use Postscript or imPRESS, and HP laser printers.

Platforms:

- o UNIX, Berkeley 4.2, System V and Xenix operating systems

Price:

- o \$995 for single user platform
- o \$1,850 for 2 - 5 user platform
- o \$2,250 for 8 - 60 user platform
- o \$4,500 for 64 - 125 user platform

## 9. Software Exoterica Corporation

Software Exoterica Corporation offers a variety of products which cover the entire SGML processing cycle. All Exoterica's products are based on a central product, the XGML Engine, which is available separately for OEM development. Unlike most other SGML vendors, Software Exoterica provides the ability to use customized DTDs with all of its products.

Software Exoterica Corporation  
383 Parkdale Avenue, Suite 406  
Ottawa, Ontario Canada K1Y 4R4  
Phone 613-722-1700  
Fax 613-722-5706

### 9.1 XGML Engine

Product Class: Application Generator

Secondary Classes: SGML/DTD Parser

The XGML Engine is a complete SGML Parser with a C subroutine library that can be used to incorporate the XGML Engine into customized applications such as database interfaces, editors, and publishing systems. The Engine features DTD compiling, interactive parsing and validation, error recovery, and conformance to both ISO 8879:1986 and Amendment 1. The XGML Engine is the basis for all other XGML products.

#### Platforms:

- o IBM mainframe (CMS)
- o VAX (VMS & Ultrix)
- o Sun (UNIX)
- o Apollo (Unix)
- o Hewlett Packard (Unix)
- o Prime (PrimeOS)
- o Macintosh (MacOS)
- o Sequent 386 (Unix)

#### Price:

- o available from vendor

### 9.2 XGML Validator

Product Class: SGML/DTD Parser

The XGML Validator is a complete SGML Parser that operates in batch mode. The Validator features DTD compiling, error recovery, and conformance to both ISO 8879:1986 and Amendment 1.



**Platforms:**

- o IBM PC 386 with 4MB (MS-DOS)
- o Sun (UNIX)

**Price:**

- o \$195 for 386 PC
- o \$295 for a Sun

### 9.3 XGML Normalizer

**Product Class:** Normalizer

**Secondary Classes:** SGML/DTD Parser

The XGML Normalizer is a complete SGML Parser that produces fully marked-up SGML from minimized input. The Normalizer features DTD compiling, error detection, error recovery, and conformance to both ISO 8879:1986 and Amendment 1. The Normalizer can parse both partial and complete documents.

**Platforms:**

- o IBM mainframe (CMS)
- o VAX (VMS & Ultrix)
- o Sun (UNIX)
- o Apollo (Unix)
- o Hewlett Packard (Unix)
- o Prime (PrimeOS)
- o Macintosh (MacOS)
- o Sequent 386 (Unix)

**Price:**

- o \$395 for 386 PC (4MB)
- o \$995 for a Sun server
- o \$1,495 for VAX server

### 9.4 XGML CheckMark

**Product Class:** DTD based SGML Editor

**Secondary Classes:** SGML/DTD Parser, Normalizer

The XGML CheckMark is a DTD based SGML editor. The editor was designed primarily for document conversion but can be used to author new documents as well. Tags can be entered from a menu or from a keyboard. A display is continuously updated with the validation status.

Validation can be done continuously while editing or while another application is running. Validation of a partial document is allowed. All validation features of SGML are supported. The CheckMark editor features DTD editing and compiling (so that any DTD can be used), error recovery, and conformance to both ISO 8879:1986 and Amendment 1.

Platforms:

- o Macintosh II, Macintosh SE/30 with 4MB memory.

Price:

- o \$495

## 9.5 XGML Translator

Product Class: Translator

Secondary Classes: Application Generator, Tagger, Normalizer, SGML/DTD Parser

The XGML Translator allows translation of SGML files to non-SGML files and vice versa. A fourth generation language, XTRAN, is used. Non-SGML documents can be converted to minimal SGML, minimized SGML or a non-SGML format (such as a formatting language). A SGML document can be parsed (using a full validating SGML/DTD parser: XGML Engine) and translated into a minimal SGML document or a document formatted in XICS, TROFF, SCRIBE, TEX, Interleaf, Frame, Rich Text Format or other formatting languages.

Platforms:

- o IBM mainframe (CMS)
- o VAX (VMS & Ultrix)
- o Sun (UNIX)
- o Apollo (Unix)
- o Hewlett Packard (Unix)
- o Prime (PrimeOS)
- o Macintosh (MacOS)
- o Sequent 386 (Unix)

Price:

- o \$19,995 for VAX or mainframe
- o \$14,995 for others

extra platforms:

- o \$5,000 for VAX or mainframe,
- o \$3,000 for others

maintenance:

- o \$7,495 comprehensive
- o \$2,245 basic

NOTE: 30 day evaluation available

## 10. Yard USA Inc.

Yard USA is a member of a European software development group (Sema Group/Yard Software Systems) specializing in management and control software. A part of this group, Sobemap, has developed several SGML products in Europe.

Yard USA Inc.  
5845 Richmond Highway, Suite 801  
Alexandria, VA 22303 USA  
Phone: 703-329-3380  
Fax: 703-960-0793

### 10.1 Write-It

Product Class: DTD Based SGML Editor

Write-It is a SGML-smart editor which can also serve as a word processor or an interface to the MARK-IT SGML Parser. Write-It features include structured document view, menu access to tags, attribute prompting, windowing, and tag minimization.

Platforms:

- o PCs (MS-DOS)

Price:

- o \$475 for single copy, discounts for multiple copies

### 10.2 MARK-IT

Product Class: SGML Parser

Yard USA is the US distributor for MARK-IT. MARK-IT is a complete SGML Parser that operates in batch mode. MARK-IT features DTD compiling, error recovery, entity declaration updating within document instances, counters to identify control and sequence of elements, creation and handling of linked document structures, linking documents to non-SGML applications such as typesetters, retroactive tagging and complete conformance to ISO 8879:1986.

Platforms:

- o Standard PC, 386 or PS/2
- o workstations/mini computers/mainframes (1- 40 terminals) (VAX-VMS,UNIX,MS-DOS)

Price:

Basic Version:

- o \$ 1,000 for Standard PC
- o \$ 2,000 for UNIX or PC Network(up to 8 seats)
- o \$ 6,000 for mainframes, mini & micro computers

Full Version:

- o \$ 5,000 for Standard PC
- o \$10,000 for UNIX or PC Network(up to 8 seats)
- o \$30,000 for mainframes, mini & micro computers

Site Licenses:

- o \$125,000 for unlimited use on single platform
- o \$160,000 for unlimited use on multiple platforms

## APPENDIX A.

### APPENDIX: Product Types by Vendor

VENDOR	PRODUCT TYPE									
	parse	DTD	normal	trans	tagger	editor	applgen	format	graphic	sys
AGFA	*	*		*	*	*		*	*	X
ArborText	*			*	*	*	*	*	*	X
Avalanche					X					
Datalogics	X	X	X	*		X	X	X	X	X
IBM	X	X		X	X	X		X	X	X
SoftQuad	X	X			X	X	*	X	*	*
Exoterica	X	X	X	X	X	X	X	X		
Yard USA	X	X				X	*			

X product type supported

\* product type not available as an individual product,  
but is part of or combination of other products from  
the vendor

## APPENDIX B.

### APPENDIX: Products and Product Types

PRODUCT TYPE										
PRODUCT	parse	DTD	normal	trans	tagger	editor	applgen	format	graphic	sys
CAPS	*	*		*	*	*		*	*	X
Publisher	*			*	*	*	*	*	*	X
FastTag					X					
IMSYS.CALS					X					
ParseStat	X	X	X	*			+			
DLPager				*				X		
IGS									X	
WriterStat	X	+	X			X				
TransDCF	X			X						
TextWrite										
OS/2	X	+				X				
TextTagger										
WorkStat					X					
ProcMaster										
TxtTaggr						X				
CALS App								X	X	
Auth/Edit	X	+				X				
MARK-IT	X	X					*			
FastTag					X					
SoftQuad P								X	*	
XGML Engin	X	X	X				X			
Validator	X	X								
Normalizer	X	X	X							
CheckMark	X	X	X			X				
Translator	X	X	X	X	X		X			
Write-It	X					X				
MARK-IT	X	X					*			

+ must purchase additional tool for this  
X product type supported  
\* part of main product type

## APPENDIX C.

### APPENDIX: Products and Platforms

Platforms						
PRODUCT	IBM PC	IBM mainframe	Macintosh	SUN	VAX	Others
CAPS				X		
Publisher				X		Apollo, HP
FastTag	X		X	X		Apollo
IMSYS.CALS	X		X	X		Apollo
ParseStat	X				X	Xenix systems
DLPager					X	
IGS					X	
WriterStat	X					
TransDCF		X				
TextWrite						
OS/2	X					
Tools	X					
TextTagger						
WorkStat	X					
ProcMaster						
TxtTaggr		X				
CALS App		X				
Auth/Edit			X	X		
Rules Bldr			X	X		
MARK-IT	X			X	X	WANG...
FastTag	X		X	X		Apollo
SoftQuad P			X	X	X	ATT, Sequent...
XGML Engin	X	X	X	X	X	:
Validator	X			X		:
Normalizer	X	X	X	X	X	check w/vendor
CheckMark			X			:
Translator	X	X	X	X	X	:
Write-It	X					
MARK-IT	X			X	X	WANG...



## APPENDIX D.

### APPENDIX: Products by Vendors

Product	Vendor
Author/Editor	SoftQuad Inc.
CAPS	ArborText, Inc.
DLPager	Datalogics Inc.
FastTag	Avalanche Development Co. or SoftQuad
IGS	Datalogics Inc.
IMSYS.CALS	Avalanche Development Company
MARK-IT	SoftQuad Inc. or Yard USA, Inc.
ParseStation	Datalogics Inc.
ProcessMaster	
TextTagger Feature	IBM Corporation
CALs Application Feature	IBM Corporation
Publisher, The	AGFA Compugraphic Division
SoftQuad Publishing System	SoftQuad Inc.
TextTagger	
WorkStation Edition	IBM Corporation
TextWrite	
OS/2 Edition	IBM Corporation
Tools OS/2 Edition	IBM Corporation
Translator DCF Edition	Yard USA, Inc.
Write-It	Datalogics Inc.
WriterStation	Software Exoterica Corp.
XGML CheckMark	Software Exoterica Corp.
XGML Engine	Software Exoterica Corp.
XGML Normalizer	Software Exoterica Corp.
XGML Translator	Software Exoterica Corp.
XGML Validator	Software Exoterica Corp.

P - Calculate and print the predicted values from the input data and the estimated model.

R - Same as P plus standard errors of the predicted and residual values, studentized residuals, and Cook's D Statistic.

The final statement we'll cover for PROC REG is the OUTPUT. OUTPUT specifies a data set into which statistics are written for each observation. The general form is:

```
PROC REG;
  MODEL ...;
  OUTPUT OUT=data-set
         statistic = names
         .
         .
         .
         statistic = names;
```

The statistics are:

P = names	Predicted values.
R = names	Residual values.
L95M = names	Lower bound of a 95% confidence interval of the mean predicted value.
U95M = names	Like L95M, but an upper bound.
L95 = names	Like L95M, but for an individual prediction.
U95 = names	Like L95, but an upper bound.
STDP = names	Standard error of the mean predicted value.
STDR = names	Standard error of the residual.
STDI = names	Standard error of the individual predicted value.
STUDENT = names	Studentized residuals; residual divided by its standard error.
COOKD = names	Cook's D influence statistic.
H = names	Leverage, $xi (X'X)^{-1}(x'i)$ .
PRESS = names	Residual for estimates dropping this observation.
RSTUDENT = names	Studentized residual with current observation deleted.
DFITS = names	Standard influence of observation on predicted values.
COVRATIO = names	Standard influence of observation on covariance of betas.

You select the names to correspond one to one with the regressor variables in the MODEL statement.

```
PROC REG;
  MODEL X = Y Z;
  OUTPUT OUT = SAVE
         P = PY PZ
         COOKD = COOKDY COOKDZ;
```

The output data set contains:

- Each variable named for a statistic (PY, PZ, COOKDY, and COOKDZ).

SAS							3
OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	RESIDUAL	STD ERR RESIDUAL	
4	C43196	1975.0	1018.5	122.6	956.5	398.2	
5	Y33142	875.0	1018.5	122.6	-143.5	398.2	
6	Y16221	950.0	1561.7	88.1694	-611.7	407.2	
7	Y48613	940.0	1561.7	88.1694	-621.7	407.2	
8	Y92114	845.0	1018.5	122.6	-173.5	398.2	
9	M47682	1200.0	1018.5	122.6	181.5	398.2	
10	M33112	1400.0	1561.7	88.1694	-161.7	407.2	
11	M28918	2000.0	2105.0	159.4	-105.0	385.0	
12	M63417	2100.0	2105.0	159.4	-4.9552	385.0	
13	R22119	900.0	1018.5	122.6	-118.5	398.2	
14	R57562	1150.0	1561.7	88.1694	-411.7	407.2	

SAS							4
OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	RESIDUAL	STD ERR RESIDUAL	
15	R41121	1180.0	1561.7	88.1694	-381.7	407.2	
16	R12221	910.0	1018.5	122.6	-108.5	398.2	
17	C11420	2115.0	1561.7	88.1694	553.3	407.2	
18	C19116	2500.0	2105.0	159.4	395.0	385.0	
19	C27226	2270.0	1561.7	88.1694	708.3	407.2	
20	C72662	2600.0	2105.0	159.4	495.0	385.0	
21	Y22131	890.0	1018.5	122.6	-128.5	398.2	
22	M14228	1900.0	1561.7	88.1694	338.3	407.2	
23	R08017	915.0	1018.5	122.6	-103.5	398.2	
24	C57251	1300.0	1018.5	122.6	281.5	398.2	

SAS							5
OBS	ID	STUDENT RESIDUAL	-2-1-0 1 2	COOK'S D			
1	Y14063	-0.4231				0.008	
2	M21364	-0.7921	*			0.054	
3	BLITZN	-0.8882	*			0.018	
4	C43196	2.4019	****			0.273	
5	Y33142	-0.3603				0.006	
6	Y16221	-1.5021	***			0.053	
7	Y48613	-1.5267	***			0.055	
8	Y92114	-0.4356				0.009	
9	M47682	0.4558				0.010	
10	M33112	-0.3971				0.004	
11	M28918	-0.2726				0.006	

## II. COURSE PROBLEMS

These problems are intended to give you some actual hands-on experience with SAS. They cover material presented in the course. Your installation may substitute its own problems in place of these.

Your first step in working these problems is to obtain a valid userid and to find out how to log on to SAS at your installation.

Using DMS or a text editor, make 10 copies of the following 10 lines so that you end up with a total of 100 lines. This data has no meaning, but doing this is an easy way for you to generate 50 observations.

Name	Age	Sex	Weight	Income
----	---	---	-----	-----
JONES, A	21	M	160	16000
SMITH, W	21	F	110	10000
JONES, A	31	M	165	20000
SMITH, W	31	F	115	35000
JONES, A	41	M	185	22000
SMITH, W	41	F	115	89000
JONES, A	51	M	215	25000
SMITH, W	51	F	115	18000
JONES, A	61	M	205	20000
SMITH, W	61	F	115	25000

Use this data in the following problems.

### PROBLEM 1

Compute the means of just the AGE variable. Request the SKEWNESS statistic.

### PROBLEM 2

Produce a frequency table with columns of SEX and rows of AGE.

### PROBLEM 3

Create a temporary SAS data set that summarizes the SUM of the WEIGHT variable by AGE. Then print this temporary data set.

### PROBLEM 4

Use PROC TABULATE to produce a table with SEX as the classification variable and AGE as the analysis variable. Make the rows of the table be SEX and the column be AGE. Let the statistic default to SUM.